DOCUMENT RESUME

ED 203 611 EC 132 896

AUTHOR Algozzine, Bob; Ysseldyke, James E.
TITLE An Analysis of Difference Score Reliabilities on Three Measures with a Sample of Low Achieving Youngsters.
INSTITUTION Minnesota Univ., Minneapolis. Inst. for Research on Learning Disabilities.
SPONS AGENCY Bureau of Education for the Handicapped (DHEW/OE), Washington, D.C.
REPORT NO IRLD-RR-37
PUB DATE Aug 80
CONTRACT 300-77-0491
NOTE 20p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Ability; *Academic Achievement; Elementary Education; *Handicap Identification; *Learning Disabilities; Low Achievement; Student Evaluation; Test Reliability
IDENTIFIERS *Ability Achievement Discrepancy

ABSTRACT

A common practice in the field of learning disabilities is analysis of ability-achievement discrepancies. The reliability of such difference scores is an important statistic in such decision making. In the study, selected ability and achievement devices were administered to a sample of 99 fourth grade low achievers (half of whom were identified as learning disabled). The reliability of various difference scores was then analyzed. In all cases, the reliabilities were moderately high. Reliabilities of differences for devices normed on the same population and differences for devices normed on different populations were comparable. (Author/DB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED203611

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

University of Minnesota

Research Report No. 37

AN ANALYSIS OF DIFFERENCE SCORE RELIABILITIES ON THREE MEASURES WITH A SAMPLE OF LOW ACHIEVING YOUNGSTERS

Bob Algozzine and James E. Ysseldyke

IRLD

Institute for Research on Learning Disabilities

EC132896

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY James Ysseldyke

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IRLD

Director: James E. Ysseldyke

Associate Director: Phyllis K. Mirkin

The Institute for Research on Learning Disabilities is supported by a contract (300-77-0491) with the Bureau of Education for the Handicapped, Department of Health, Education, and Welfare, U.S. Office of Education, through Title VI-G of Public Law 91-230. Institute investigators are conducting research on the assessment/decision-making/intervention process as it relates to learning disabled children. Research activities are organized into eight major areas:

I. Adequacy of Norm-Referenced Data for Prediction of Success

II. Computer Simulation Research on the Assessment/Decision-making/Intervention Process

III. Comparative Research on Children Labeled LD and Children Failing Academically but not Labeled LD

IV. Surveys on In-the-Field Assessment, Decision Making, and Intervention

V. Ethological Research on Placement Team Decision Making

VI. Bias Following Assessment

VII. Reliability and Validity of Formative Evaluation Procedures

VIII. Data-Utilization Systems in Instructional Programming

Additional information on these research areas may be obtained by writing to the Editor at the Institute.

The research reported herein was conducted under government sponsorship. Contractors are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official position of the Bureau of Education for the Handicapped.

Research Report No. 37

AN ANALYSIS OF DIFFERENCE SCORE RELIABILITIES ON THREE MEASURES WITH A SAMPLE OF LOW ACHIEVING YOUNGSTERS

Bob Algozzine and James E. Ysseldyke

Institute for Research on Learning Disabilities

University of Minnesota

August, 1980

Abstract

A common practice in the field of learning disabilities is analysis of ability-achievement discrepancies. The reliability of such difference scores is an important statistic in such decision making. In this study, selected ability and achievement devices were administered to a sample of 99 low achievers; the reliability of various difference scores was analyzed. In all cases, the reliabilities were moderately high. Reliabilities of differences for devices normed on the same population and differences for devices normed on different populations were comparable. These results are discussed in light of current psychometric practices.

An Analysis of Difference Score Reliabilities on Three Measures with a Sample of Low Achieving Youngsters

The U. S. Office of Education, in its final rules and regulations on Procedures for Evaluating Specific Learning Disabilities (Federal Register, 1977), established criteria to be used in identifying students with specific learning disabilities. To be identified as learning disabled, team members must demonstrate that a child (1) does not achieve commensurate with his or her age and ability levels in one or more of seven areas when provided with learning experiences appropriate for the child's age and ability levels, and (2) the child has a "severe discrepancy" between achievement and intellectual ability in one or more of seven areas related to communication skills and mathematical abilities (oral expression, listening comprehension, written expression, basic reading skill, reading comprehension, mathematics calculation, and mathematical reasoning).

The rules and regulations do not specify the magnitude of a discrepancy needed to be considered "severe." Yet, diagnostic personnel regularly use deficit scores to identify the learning disabled. Considerable variation exists in the approaches used to identify "severe discrepancies" (cf. Volume 2, issue 4 of Learning Disability Quarterly, 1979). Many approaches fail to consider the reliability of difference scores, a practice strongly recommended by Salvia and Ysseldyke (1978).

The purpose of this investigation was to ascertain the reliabilities of difference scores obtained through administration of the Woodcock-Johnson Psycho-Educational Battery, the Revised Wechsler Intelligence Scale for Children (WISC-R), and the Peabody Individual Achievement Test

(PIAT). In addition, we investigated the extent to which reliabilities of differences were a function of the comparability of groups on whom the tests were standardized.

Method

Subjects

Ninety-nine fourth graders from nine different school districts participated in this study. Fifty of the students had been labeled "learning disabled" by their respective school districts. Criteria used to identify these students differed in the different school districts. The students were low achievers, as suggested by their average total score on the Peabody Individual Achievement Test ($\bar{X}$ = 91.9, SD = 8.78).

Forty-nine of the students were fourth graders who had not been identified as learning disabled, but who were selected as low achievers on the basis of having scored below the 25th percentile on the Iowa Tests of Basic Skills administered during the fall of the year in which these data were collected.

The ages, sex distribution, parental marital status, family socio-economic status, and family income of the two groups were not significantly different. Ysseldyke, Algozzine, Shinn, and McGue (in press) reported that there were no differences of practical significance in the performance of the two groups on 49 psychometric measures in the domains of intelligence, achievement, visual-motor functioning, self-concept, and problem behavior. For purposes of this investigation, the two groups were combined and labeled as low achievers. Demographic characteristics of the total group of 99 subjects are reported in Table 1.

Insert Table 1 about here

Procedure

Each subject was administered a battery of tests during participation in a larger study. All testing was completed by qualified psychometricians and occurred during the same period of time (i.e., January to May). Demographic information was collected from the parent(s) of the children and a behavior rating scale was completed by their current teacher.

The test battery included the Wechsler Intelligence Scale for Children-Revised (WISC-R), the Peabody Individual Achievement Test (PIAT), selected subtests of the Stanford Achievement Test (SAT), the Bender Visual-Motor Gestalt Test (BVMGT), the Developmental Test of Visual-Motor Integration (CTVMI), the Piers-Harris Self-Concept Scale, and the Peterson-Quay Behavior Problem Checklist. Descriptions of each of these devices, including information on their technical adequacy, are included in Salvia and Ysseldyke (1978). The battery was selected as one including those devices commonly used with LD youngsters.

Additionally, selected subtests of the Woodcock-Johnson (WJ) Psycho-Educational Battery were administered to each student. The Battery includes 27 subtests designed to measure cognitive abilities, scholastic aptitudes, achievement, and interests. Woodcock (1978) suggests that the Battery be used to identify students with "special problems or disabilities" through analysis of discrepancies between aptitude and achievement, a procedure requiring the use of difference scores.

We used all 12 subtests of Part One: Tests of Cognitive Ability (Picture Vocabulary, Spatial Relations, Memory for Sentences, Visual-

Auditory Learning, Blending, Quantitative Concepts, Visual Matching, Antonyms-Synonyms, Analysis-Synthesis, Numbers Reversed, Concept Formation, and Analogies), as well as seven of the 10 Tests of Achievement subtests (Letter-Word Identification, Word Attack, Passage Comprehension, Calculation, Applied Problems, Dictation, and Proofing). No subtests from Part Three (Tests of Interest Level) were administered.

Data Analyses

Correlations between selected "ability," "aptitude," and "achievement" measures were obtained as were reliabilities of differences between them. Actual reliabilities for each device were available in test manuals; those for fifth grade children were used in each instance. Specifically, relationships between the WISC-R, WJ Broad Cognitive cluster scores, and WJ Aptitude cluster scores (i.e., ability measures) and the PIAT scores and WJ Achievement cluster scores (i.e., achievement measures) were analyzed.

Results

The correlations between scores from the various psychometric devices are presented in Table 2; the reported reliability for each subtest or test (taken from the manuals) appears in parentheses after the name of each device. In general, the reliabilities were acceptable for most types of decision making and the correlations between scores were moderately high. The reliabilities of the differences between scores on various devices are presented in Table 3; most of these are lower than either of the reliabilities of the tests from which they were derived.

Insert Tables 2 and 3 about here

The reliabilities of differences between the WJ ability and achievement measures are quite similar and are not different from those between the WISC-R ability and WJ achievement discrepancies. For the WJ achievement clusters, it should be noted that, in some cases, similar subtests are grouped together to form the different achievement cluster scores. For example, Reading is comprised of Letter-Word Identification, Word Attack, and Passage Comprehension subtest scores and Skills is composed of Letter-Word Identification, Applied Problems, and Dictation subtest scores. Considerably more overlap exists among the subtests which comprise the aptitude clusters; Antonyms-Synonyms scores are included in all aptitude clusters, Analogies scores are included in Reading and Knowledge Aptitude cluster scores, Quantitative Concepts scores are included in Written Language and Knowledge Aptitude Cluster scores, and Visual Matching scores are included in Mathematics and Written Language Aptitude cluster scores. Additionally, the Broad Cognitive Ability cluster score is comprised of all "cognitive ability" subtest scores.

The reliabilities of differences between the PIAT achievement subtests and each of the ability measures were also relatively similar. It seems, from this analysis, that no distinct advantage relative to reliability is evidenced in differences obtained through use of the WISC-R, WJ Battery, and PIAT. In all cases, the differences have a similar degree of reliability; this in no way suggests that the use of those differences is a valid or recommended diagnostic technique.

Discussion

A common and often recommended practice for decision making in the field of learning disabilities is the identification of discrepancies

within a child's test performance. The importance of this practice is underlined by regulations present in Public Law 94-142 (Education for All Handicapped Children Act). This research was designed to evaluate the reliability of difference (i.e., discrepancy) scores for selected psychometric devices administered to a sample of underachieving youngsters.

An analysis of the results indicated that comparable statistics were obtained when different devices were administered. In all cases, the reliability of difference scores was moderately high (e.g., range = 0.52 to 0.93). Similarly, differences in these reliabilities did not seem to be a function of the devices on which they were based.

In discussing the use of "profile analysis" and "analysis of difference scores," Salvia and Ysseldyke (1978) suggested that these tasks are difficult because typically the tests used to define ability-achievement discrepancies are normed on different samples and correlations between them generally are not available. The problems inherent in decision-making devices normed on different samples have not been addressed by this research; however, we have obtained correlations between tests often used in making decisions about "discrepancies." The reliabilities of differences obtained with tests of different norms (e.g., WISC-R and PIAT) were similar to those obtained for a devices with one normative population (e.g., WJ). The extent to which these data are useful to decision makers remains the subject of both applied and basic research.

References

Federal Register. U. S. Office of Education, Bureau of Education for the Handicapped, 1977, 42, 65082-65085.

Salvia, J., & Ysseldyke, J. E. Assessment in special and remedial education. Boston: Houghton-Mifflin, 1978.

Woodcock, R. W. Development and standardization of the Woodcock-Johnson Psycho-Educational Battery. Boston: Teaching Resources, 1978.

Ysseldyke, J. E., Algozzine, B., Shinn, M., & McGue, M. Similarities and differences between underachievers and students labeled learning disabled. Journal of Special Education, in press.

Footnote

Bob Algozzine is also Associate Professor of Special Education at the University of Florida, Gainesville.

The authors wish to acknowledge the contributions of Mark Shinn and Matthew McGue to data compilation and analysis.

Table 1

Demographic Characteristics of 99 Low Achievers

Child's Sex		Child's Age (months)		Parental Marital Status		Family SES		Family Income	
Male	Female	$\bar{X}$	SD	Married	Unmarried	$\bar{X}$	SD	$\bar{X}$	SD
75	24	121.50	4.52	54	17	50.91	23.91	$21,138	$10,752

Table 2

Correlations between Selected Psychometric Devices

Ability Scores	Achievement Scores									
	PIAT Total (.89)	PIAT Mathematics (.73)	PIAT Reading Rec (.89)	PIAT Reading Comp (.89)	PIAT Spelling (.53)	PIAT General Info (.88)	WJ Reading (.95)	WJ Mathematics (.91)	WJ Written Language (.94)	WJ Skills (.95)
WISC-R Full Scale IQ (.95)	.44	.45	.19	.34	.06	.50	.36	.52	.35	.43
WISC-R Verbal IQ (.93)	.52	.49	.27	.39	.14	.56	.41	.58	.41	.51
WISC-R Performance IQ (.89)	.21	.26	.05	.18	-.07	.27	.19	.31	.17	.21
WJ Broad Cognitive Cluster Score (.96)	.52	.53	.29	.48	.15	.46	.37	.62	.42	.56
WJ Math Aptitude Cluster Score (.86)	.49	.27	.44	.54	.13	.45	.38	.51	.40	.52
WJ Reading Aptitude Cluster Score (.93)	.51	.38	.43	.44	.16	.45	.53	.48	.49	.56
WJ Writing Aptitude Cluster Score (.90)	.58	.35	.52	.55	.30	.46	.42	.67	.45	.65
WJ Knowledge Aptitude Cluster Score (.93)	.61	.36	.52	.53	.20	.60	.45	.61	.49	.62

Note. Numbers in parentheses are reliabilities for each device. Number of subjects = 99.

Table 3

Reliability of Difference Scores for Selected Psychometric Devices

Ability Scores	Achievement Scores									
	PIAT Total (.89)	PIAT Mathematics (.73)	PIAT Reading Rec (.89)	PIAT Reading Comp (.89)	PIAT Spelling (.53)	PIAT General Info (.88)	WJ Reading (.95)	WJ Mathematics (.91)	WJ Written Language (.94)	WJ Skills (.95)
WISC-R Full Scale IQ (.95)	.86	.71	.90	.69	.73	.83	.92	.85	.92	.91
WISC-R Verbal IQ (.93)	.81	.68	.88	.65	.69	.29	.90	.81	.89	.88
WISC-R Performance IQ (.89)	.86	.74	.88	.71	.73	.84	.90	.86	.90	.90
WJ Broad Cognitive Cluster Score (.96)	.84	.67	.90	.62	.70	.85	.93	.83	.91	.90
WJ Math Aptitude Cluster Score (.86)	.76	.83	.55	.56	.65	.76	.85	.76	.83	.80
WJ Reading Aptitude Cluster Score (.93)	.82	.85	.62	.70	.68	.83	.87	.85	.87	.87
WJ Writing Aptitude Cluster Score (.90)	.75	.84	.52	.59	.59	.80	.87	.71	.86	.78
WJ Knowledge Aptitude Cluster Score (.93)	.77	.86	.55	.64	.66	.76	.89	.80	.87	.84

Note. Numbers in parentheses are reliabilities for each device (taken from the respective test manuals). Number of subjects = 99.

PUBLICATIONS

Institute for Research on Learning Disabilities
University of Minnesota

The Institute is not funded for the distribution of its publications. Publications may be obtained for $3.00 per document, a fee designed to cover printing and postage costs. Only checks and money orders payable to the University of Minnesota can be accepted. All orders must be prepaid.

Requests should be directed to: Editor, IRLD, 350 Elliott Hall; 75 East River Road, University of Minnesota, Minneapolis, MN 55455.

Ysseldyke, J. E. Assessing the learning disabled youngster: The state of the art (Research Report No. 1). November, 1977.

Ysseldyke, J. E., & Regan, R. R. Nondiscriminatory assessment and decision making (Monograph No. 7). February, 1979.

Foster, G., Algozzine, B., & Ysseldyke, J. Susceptibility to stereotypic bias (Research Report No. 3). March, 1979.

Algozzine, B. An analysis of the disturbingness and acceptability of behaviors as a function of diagnostic label (Research Report No. 4). March, 1979.

Algozzine, B., & McGraw, K. Diagnostic testing in mathematics: An extension of the PIAT? (Research Report No. 5). March, 1979.

Deno, S. L. A direct observation approach to measuring classroom behavior: Procedures and application (Research Report No. 6). April, 1979.

Ysseldyke, J. E., & Mirkin, P. K. Proceedings of the Minnesota round-table conference on assessment of learning disabled children (Monograph No. 8). April, 1979.

Somwaru, J. P. A new approach to the assessment of learning disabilities (Monograph No. 9). April, 1979.

Algozzine, B., Forgnone, C., Mercer, C. D., & Trifiletti, J. J. Toward defining discrepancies for specific learning disabilities: An analysis and alternatives (Research Report No. 7). June, 1979.

Algozzine, B. The disturbing child: A validation report (Research Report No. 8). June, 1979.

Note: Monographs No. 1 - 6 and Research Report No. 2 are not available for distribution. These documents were part of the Institute's 1979-1980 continuation proposal, and/or are out of print.

Ysseldyke, J. E., Algozzine, B., Regan, R., & Potter, M. Technical adequacy of tests used by professionals in simulated decision making (Research Report No. 9). July, 1979.

Jenkins, J. R., Deno, S. L., & Mirkin, P. K. Measuring pupil progress toward the least restrictive environment (Monograph No. 10). August, 1979.

Mirkin, P. K., & Deno, S. L. Formative evaluation in the classroom: An approach to improving instruction (Research Report No. 10). August, 1979.

Thurlow, M. L., & Ysseldyke, J. E. Current assessment and decision-making practices in model programs for the learning disabled (Research Report No. 11). August, 1979.

Deno, S. L., Chiang, B., Tindal, G., & Blackburn, M. Experimental analysis of program components: An approach to research in CSDC's (Research Report No. 12). August, 1979.

Ysseldyke, J. E., Algozzine, B., Shinn, M., & McGue, M. Similarities and differences between underachievers and students labeled learning disabled: Identical twins with different mothers (Research Report No. 13). September, 1979.

Ysseldyke, J., & Algozzine, R. Perspectives on assessment of learning disabled students (Monograph No. 11). October, 1979.

Poland, S. F., Ysseldyke, J. E., Thurlow, M. L., & Mirkin, P. K. Current assessment and decision-making practices in school settings as reported by directors of special education (Research Report No. 14). November, 1979.

McGue, M., Shinn, M., & Ysseldyke, J. Validity of the Woodcock-Johnson psycho-educational battery with learning disabled students (Research Report No. 15). November, 1979.

Deno, S., Mirkin, P., & Shinn, M. Behavioral perspectives on the assessment of learning disabled children (Monograph No. 12). November, 1979.

Sutherland, J. H., Algozzine, B., Ysseldyke, J. E., & Young, S. What can I say after I say LD? (Research Report No. 16). December, 1979.

Deno, S. L., & Mirkin, P. K. Data-based IEP development: An approach to substantive compliance (Monograph No. 13). December, 1979.

Ysseldyke, J., Algozzine, B., Regan, R., & McGue, M. The influence of test scores and naturally-occurring pupil characteristics on psycho-educational decision making with children (Research Report No. 17). December, 1979.

Algozzine, B., & Ysseldyke, J. E. Decision makers' prediction of students' academic difficulties as a function of referral information (Research Report No. 18). December, 1979.

Ysseldyke, J. E., & Algozzine, B. Diagnostic classification decisions as a function of referral information (Research Report No. 19). January, 1980.

Deno, S. L., Mirkin, P. K., Chiang, B., & Lowry, L. Relationships among simple measures of reading and performance on standardized achievement tests (Research Report No. 20). January, 1980.

Deno, S. L., Mirkin, P. K., Lowry, L., & Kuehnle, K. Relationships among simple measures of spelling and performance on standardized achievement tests (Research Report No. 21). January, 1980.

Deno, S. L., Mirkin, P. K., & Marston, D. Relationships among simple measures of written expression and performance on standardized achievement tests (Research Report No. 22). January, 1980.

Mirkin, P. K., Deno, S. L., Tindal, G., & Kuehnle, K. Formative evaluation: Continued development of data utilization systems (Research Report No. 23). January, 1980.

Deno, S. L., Mirkin, P. K., Robinson, S., & Evans, P. Relationships among classroom observations of social adjustment and sociometric rating scales (Research Report No. 24). January, 1980.

Thurlow, M. L., & Ysseldyke, J. E. Factors influential on the psychoeducational decisions reached by teams of educators (Research Report No. 25). February, 1980.

Ysseldyke, J. E., & Algozzine, B. Diagnostic decision making in individuals susceptible to biasing information presented in the referral case folder (Research Report No. 26). March, 1980.

Thurlow, M. L., & Greener, J. W. Preliminary evidence on information considered useful in instructional planning (Research Report No. 27). March, 1980.

Ysseldyke, J. E., Regan, R. R., & Schwartz, S. Z. The use of technically adequate tests in psychoeducational decision making (Research Report No. 28). April, 1980.

Richey, L., Potter, M., & Ysseldyke, J. Teachers' expectations for the siblings of learning disabled and non-learning disabled students: A pilot study (Research Report No. 29). May, 1980.

Thurlow, M. L., & Ysseldyke, J. E. Instructional planning: Information collected by school psychologists vs. information considered useful by teachers (Research Report No. 30). June, 1980.

Algozzine, B., Webber, J., Campbell, M., Moore, S., & Gilliam, J. Classroom decision making as a function of diagnostic labels and perceived competence (Research Report No. 31). June, 1980.

Ysseldyke, J. E., Algozzine, B., Regan, R. R., Potter, M., Richey, L., & Thurlow, M. L. Psychoeducational assessment and decision making: A computer-simulated investigation (Research Report No. 32). July, 1980.

Ysseldyke, J. E., Algozzine, B., Regan, R. R., Potter, M., & Richey, L. Psychoeducational assessment and decision making: Individual case studies (Research Report No. 33). July, 1980.

Ysseldyke, J. E., Algozzine, B., Regan, R., Potter, M., & Richey, L. Technical supplement for computer-simulated investigations of the psychoeducational assessment and decision-making process (Research Report No. 34). July, 1980.

Algozzine, B., Stevens, L., Costello, C., Beattie, J., & Schmid, R. Classroom perspectives of LD and other special education teachers (Research Report No. 35). July, 1980.

Algozzine, B., Siders, J., Siders, J., & Beattie, J. Using assessment information to plan reading instructional programs: Error analysis and word attack skills (Monograph No. 14). July, 1980.

Ysseldyke, J., Shinn, M., & Epps, S. A comparison of the WISC-R and the Woodcock-Johnson Tests of Cognitive Ability (Research Report No. 36). July, 1980.

Algozzine, B., & Ysseldyke, J. E. An analysis of difference score reliabilities on three measures with a sample of low achieving youngsters (Research Report No. 37). August, 1980.

Shinn, M., Algozzine, B., Marston, D., & Ysseldyke, J. A theoretical analysis of the performance of learning disabled students on the Woodcock-Johnson Psycho-Educational Battery (Research Report No. 38). August, 1980.